Statistical Analysis of Computational Docking of Large Compound Data Bases to Distinct Protein Binding Sites

JEFFREY W. GODDEN,[1] FLORENCE L. STAHURA,[1] JÜRGEN BAJORATH[2]

[1] *MDS Panlabs, Computational Chemistry & Informatics, 11804 North Creek Parkway South, Bothell, Washington 98011-8805*

[2] *MDS Panlabs and Department of Biological Structure, University of Washington, Seattle, Washington 98195*

Received 6 January 1998; accepted 28 June 1999

ABSTRACT: The results of 16 docking simulations with rigid receptor sites and flexible ligands (~ 60,000 compounds in each case) are statistically analyzed and compared. Different combinations of binding sites, scoring functions, and compound collections are used in these calculations. The docking scores are not randomly distributed over the scoring range; they follow Gaussian distributions (regardless of the binding sites), scoring functions, or screened compounds. If the docking sites are small, the Gaussian distributions are positively skewed. Peaks of the Gaussian distributions are populated with compounds having similar scores but different sizes and binding modes. These findings have implications for compound selection via computational docking. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 1634–1643, 1999

Keywords: computational docking; statistical analysis; score distributions; skewness; protein–ligand interactions

Introduction

Computational docking techniques were developed to predict and analyze protein–ligand and protein–protein interactions.[1,2] A variety of docking algorithms and computational implementations are available.[3–6] Automated docking of ligand data bases to proteins has become a widely used approach to aid in the identification of inhibitors.[4,7]

Protein–ligand docking requires the definition of the receptor site and is typically carried out by shape matching[8,9] of the receptor site and rigid or flexible ligands.[9,10] Ligand flexibility can be taken into account explicitly via flexible search[10] or im-

Correspondence to: J. Bajorath; e-mail: jbajorath@nce-mail.com

CCC 0192-8651 / 99 / 151634-10

plicitly using pregenerated multiple conformations of ligands.[11] Limited flexibility of receptor sites is also considered.[12]

Typically, predicted protein–ligand complexes are evaluated using scoring functions that attempt to capture the degree of surface and/or chemical complementarity between the receptor and ligand.[4,5] In contact scoring only the surface complementarity is evaluated. Energy scoring typically uses force field energies calculated *in vacuo* as an approximate measure of protein–ligand interactions.

The focal point of many protein–ligand docking studies is to aid in structure-based design (i.e., identification of candidate inhibitors by virtual screening of compound data bases).[7] Thus, the computer experiment is primarily carried out to identify a number of ''hits'' for experimental analysis. How is this done? Typically, a number of well-scoring compounds are selected and further analyzed (often involving subjective criteria), and the most promising candidates are selected.[13,14]

In an actual screening experiment most compounds in a data base do not bind to a given target; otherwise, ''specific'' inhibitors would not exist. However, in a computer experiment all docked compounds ''bind.'' Therefore, evaluating and scoring docked compounds becomes a critical task in docking simulations. Contemporary scoring functions, as described above, can only roughly estimate complex interactions and other effects [e.g., (de)solvation] that discriminate between binding and nonbinding. Therefore, the approximate nature of scoring functions and the difficulties in accounting for conformational effects upon binding are major problems of automated docking.[5,11]

It was the uncertainty of what defines a hit in a computer experiment (e.g., the docking scores) that prompted us to investigate a novel aspect of automated docking: the range and distribution of docking scores obtained for screening of large compound collections. Flexible ligand docking with anchor fragments[10,15] as implemented in the DOCK program[9,16] was used to dock a total of ~ 120,000 diverse compounds to three target enzymes: human carbonic anhydrase II (CA),[17] cAMP-dependent Ser/Thr kinase (PK),[18] and $p56^{lck}$ tyrosine kinase (LK).[19]

The nature of the binding site in CA differs significantly from PK and LK. The targeted ATP binding sites in these two kinases are similar, yet they have different volumes when transformed into docking sites. Different scoring functions were applied and the resulting scores of all compounds were analyzed, including those obtained for experimentally determined enzyme–inhibitor complexes. In all simulations the distributions of scores were Gaussian-like, and peak regions corresponded to diverse compounds with different binding modes. Positive skewness of the distributions was observed when the receptor sites were small, which is consistent with the idea that small docking sites are less permissive to accommodate random binding.

Materials and Methods

Structural manipulations were carried out with MOE (Molecular Operating Environment, Chemical Computing Group Inc., Montreal) and representations were generated with InsightII (MSI, San Diego, CA). X-ray structures of CA, LK, and PK were obtained from the Brookhaven Protein Data Bank.[20] All docking calculations were carried out with DOCK 4.0.1[16] and flexible ligands using anchored search.[10,15] Atomic partial charges on the protein atoms were calculated with AMBER.[21] The CA active site zinc ion was assigned a charge of +2.0. The Connolly surface was calculated with the program MS[22] using a surface density of five surface points per $Å^2$ and a probe radius of 1.4 Å. Sphere models were calculated with the DOCK-associated program SPHGEN.

The PK binding site for the docking experiments was created using PDB entry 1YRD after removal of the inhibitor 1-(5-isoquinolinesulfonyl)-2-methyl- piperazine (PIP).[18] The pdb entry 3LCK was used to build the LK binding site.[19] For both kinases there were 28 residues selected. These residues corresponded to those within 9 Å of the first phosphorus atom in the structure of the ternary complex of PK with MgATP and peptide inhibitor (pdb entry 1ATP).[23] Sixty-two spheres and 61 spheres were selected for PK and LK, respectively, corresponding to a volume of 660 $Å^3$ for the PK site and 1370 $Å^3$ for the LK site. Smaller PK and LK docking sites were created by reducing the number of spheres to 24 spheres for PK, corresponding to 450 $Å^3$, and to 15 spheres for LK, corresponding to 730 $Å^3$.

For CA the PDB entry 1AM6[17] was selected as a representative structure. The bound inhibitor acetohydroxamate was removed. The site was isolated for DOCK calculations by deleting all residues that were at least 10 Å from the active site

zinc atom. For CA there were 32 spheres generated that covered 500 Å^3.

For compound screening two small molecular data bases were used, Optiverse™, a diverse combinatorial library,[24] and Maybridge (Maybridge 97, Daylight Chemical Information Systems, Irvine, CA), each containing ~ 60,000 compounds with an average molecular weight of 350. Hydrogen atoms and partial atomic charges[25] were added to all compounds. Potential 3-dimensional conformations were generated using MOE. Each compound was docked using two scoring functions, contact scoring (shape complementarity) and energy scoring (AMBER force field energy).

Scores were calculated for X-ray structures of CA in complex with (4*s-trans*)-4-(ethylamino)-5,6-dihydro-6-methyl-4thieno(2,3-*b*) thiopyran-2-sulfonamide-7,7-dioxide (ETS), (4*s*-trans)-4-(methylamino)-5, 6-dihydro-6-methyl-4thieno(2, 3-*b*) thiopyran-2-sulfonamide-7,7-dioxide (MTS), and ethylaminocarbonyl-benzene-sulfonamide (EAS). Scores were also calculated for the X-ray structure of PIP bound to PK[18] and for ATP after superposition of the PK docking site on the X-ray structure of ternary complex 1ATP (see above).

For statistical analysis the DOCK scores were assigned to 200 bins of equal size covering the entire range of scores. The number of bins was empirically determined to minimize the number of unpopulated bins. Histograms were plotted with the number of molecules per bin for each bin on the abscissa. For example, the range of contact scores for the Maybridge data base screened on PK spanned −200 to −20 kcal/mol. Therefore, a single bin contained the number of data base molecules with scores within 0.9 kcal/mol. The automated extraction of scores from DOCK output files, the binning of the data, and calculation of mean, mode, standard deviation, and skewness (the third moment) of the distribution was performed using perl scripts. Skewness is defined as

$$\text{skew}\,(x_1 \cdots x_N) = 1/N \sum_{J=1}^{N} \left[(x_j - x_{\text{obs}})/\sigma\right]^3,$$

where $\sigma = \sigma(x_1 \ldots x_N)$ is the standard deviation of the distribution, N is the number of data points, and x_{obs} is the most frequently observed value within each distribution. Because a skewed distribution is poorly characterized by a mean central value, all calculations that required a central value used the conveniently available mode (most frequently observed value).

Results and Discussion

Three target enzymes were selected to present different docking situations. Compared to LK and PK, CA has a very different binding site and catalytic mechanism. Docking sites were defined based on the X-ray structures of PK and CA in complex with inhibitors and based on uncomplexed LK. The defined docking sites in PK and LK correspond to the ATP binding site that is largely conserved in protein kinases. Figure 1 shows the ATP binding site of PK with a bound inhibitor. For both PK and LK there were 28 residues selected to represent the binding site. However, the generated sphere model representing the docking site in LK is approximately twice as large as in PK. Furthermore, the docking site in LK is approximately 4 times the volume of the docking site in CA, which is centered on the active site zinc ion. The volume difference between the LK and PK docking sites results from inhibitor binding to PK, which constricts the size of the binding site.[18,19] This is illustrated in Figure 2, which compares the docking sites in LK and PK. Thus, the selected docking sites have very different sizes and shapes. Moreover, ligand binding to CA involves coordination of a zinc ion in the active site, which is difficult, if not impossible, to account for in molecular mechanics- or shape-based calculations. Therefore, CA is a ''difficult'' docking target that was deliberately chosen along with more suitable targets (e.g., kinases) to test the generality of score distributions.

Test calculations were carried out for PK and CA to reproduce the experimental binding mode of known inhibitors (PIP and ETS, respectively; see Materials and Methods) that were added to the compound data bases in their binding conformations. For PK the experimental binding mode of the inhibitor could be well reproduced [all atom root mean square (RMS) deviation of ~ 0.7 Å, taking positional differences into account]. By contrast, this was not possible for CA. As mentioned above, known sulfonamide inhibitors directly coordinate the active site zinc in CA, and this precise geometric arrangement could not be reproduced, as anticipated, by docking and scoring based on shape complementarity and/or force field energies. Docked orientations were very different from the experiment (''best'' RMS deviations from the experiment were ~ 16 Å). Thus, PK is a target

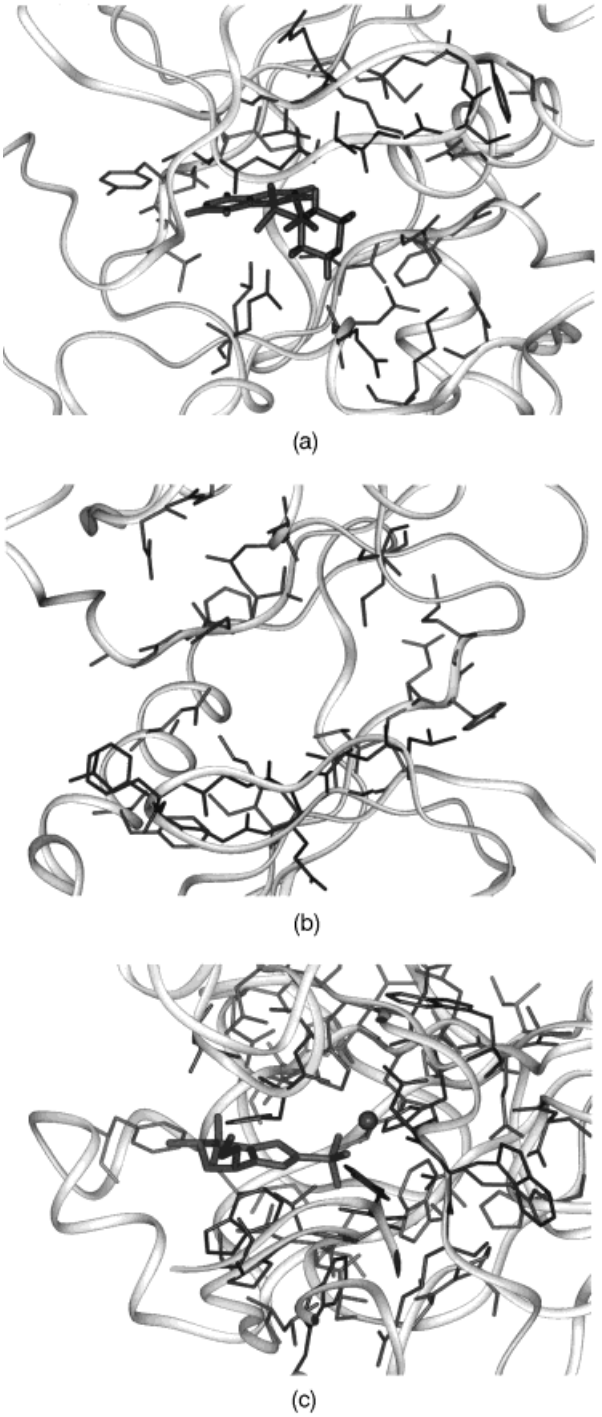

FIGURE 1. Docking sites. In (a), the binding site of PK with bound inhibitor PIP is shown. PK is represented as a ribbon. All residues selected to generate the docking site are shown. In (b), the empty binding site of LK is depicted. In (c), the binding site of CA is shown with inhibitor ETS.

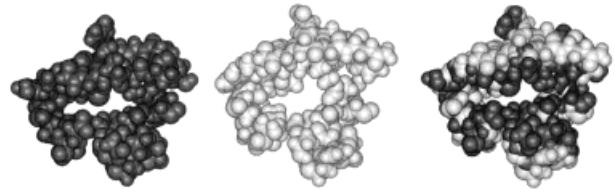

FIGURE 2. Comparison of docking sites in PK and LK. Residues forming the docking sites are shown with a solid van der Waals surface. From the left to the right, PK (dark), LK (light), and the superposition of both are shown. The figure illustrates the constriction of the PK site upon inhibitor binding.

enzyme for which experimentally determined structures could be reproduced whereas CA is not.

Two different compound data bases were selected for docking, Optiverse™, a diverse screening library,[24] and the Maybridge data base, which contains many compounds commonly used in medicinal chemistry. Compounds were docked taking (fragment based) ligand flexibility into account. Although computationally expensive, this allows the evaluation of each ligand without depending on a precalculated data base conformation. Two scoring functions were applied, an intermolecular contact score and an energy score. The contact score estimates shape the complementarity of the receptor and ligand and only considers heavy atom contacts and van der Waals overlap. The energy score includes a molecular mechanics force field estimate of electrostatic and van der Waals interactions using a united atom model.[21] Other scoring functions are available to analyze protein–ligand interactions, for example, empirical free energy functions,[26] but they have not yet been implemented in DOCK. Furthermore, scoring approaches that depend on extensive hydrophobicity calculations were not applied in this study because they are computationally too expensive for rapid screening of large compound data bases.

A total of 16 docking simulations was carried out. In each calculation ~ 60,000 compounds were docked. DOCK parameters are summarized in Table I. The results were statistically analyzed using a histogram method. In this analysis, the range of docking scores was determined, small energy bins were defined to optimally cover the scoring range (see Materials and Methods), and the number of compounds in each bin was determined. The results of the statistical analysis for docking to each target are summarized in Table II.

TABLE I.
DOCK Parameters.

Flexible_ligand	Yes	Bump_filter	Yes
Orient_ligand	Yes	Bump_maximum	0
Score_ligand	Yes	Contact_score	Yes
Minimize_ligand	No	Contact_cutoff_distance	4.5
Multiple_ligands	Yes	Contact_clash_overlap	0.75
Parallel_jobs	No	Contact_clash_penalty	50
Anchor_search	Yes	Chemical_score	No
Multiple_anchors	Yes	Energy_score	Yes
Anchor_size	10	Energy_cutoff_distance	10
Torsion_drive	Yes	Distance_dielectric	Yes
Clash_overlap	0.5	Dielectric_factor	4
Configurations_per_cycle	25	Attractive_exponent	6
Flexible_bond_maximum	10	Repulsive_exponent	12
Match_receptor_sites	Yes	Atom_model	u
Random_search	No	vdw_scale	1
Automated_matching	Yes	Electrostatic_scale	1
Maximum_orientations	500	Ligands_maximum	100,000
Write_configurations	No	Initial_skip	0
Intramolecular_score	Yes	Interval_skip	0
Intermolecular_score	Yes	Heavy_atoms_minimum	0
Gridded_score	Yes	Heavy_atoms_maximum	200
Grid_version	4	Rank_ligands	No

The distribution of scores was similar for LK (Fig. 3), PK (Fig. 4), and CA (Fig. 5), irrespective of the compound data base and scoring function. Figures 3a-d, 4a-d, and 5a-d show representative distributions. For all targets the distributions conform well to simple unimodal (skewed) Gaussian distributions. Such a distribution divides a population into two extremes and a central typical mass. In each case this central peak section contains the bulk of the docked compounds. The major difference between LK, PK, and CA is the skewness of the distributions. The skewness is statistically insignificant for LK, which has the largest docking site, but positive for both PK and CA (further discussed below).

We next analyzed the molecular weights and binding modes of compounds that populate the peak sections of the Gaussian distributions (Fig. 6). Compounds in the peak sections have different molecular weights (which mirror the weight distribution of the compound data base). Furthermore, the binding modes of compounds in the peak sections vary greatly (Fig. 7). Thus, the peak segments of the distributions correspond to many compounds having similar scores but different sizes and binding modes. Thus, in the computer simulation this could be interpreted as a scoring range corresponding to ''nonspecific binding.''

As reference points the scores were calculated for X-ray structures of PK in complex with its cofactor ATP or the inhibitor PIP (see Materials and Methods) and for CA in complex with the inhibitors ETS, MTS, and EAS (see Materials and Methods). The scoring positions of these compounds are indicated in Figures 4 and 5. For both PK and CA the ligands score within 1 standard deviation of the peaks of the distributions. That experimentally determined inhibitors score close to the nonspecific binding range illustrates the shortcomings of scoring functions. Because the best scoring compounds are usually considered hits in the docking simulation, the positions of known ligands in the distribution illustrate that this selection criterion is problematic.

The skewness of the distributions was statistically zero for LK but positive for PK and CA (Table II). Positively skewed distributions have a sharper peak and more compounds in the ''worse than nonspecific'' scoring range. The binding sites are similar in PK and LK but the volume of the

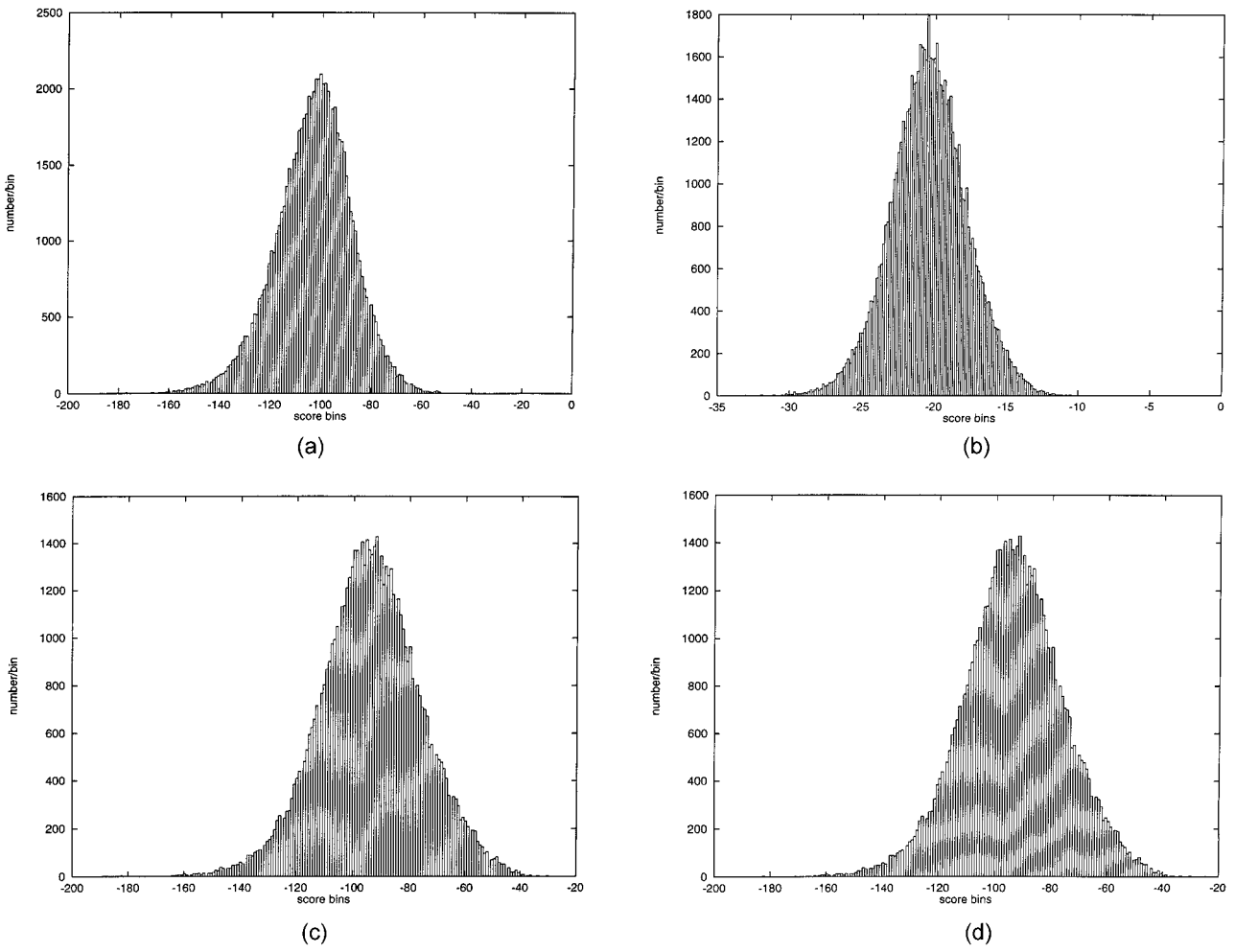

FIGURE 3. Histograms of LK docking scores. (a) Contact scores with Optiverse™ compounds, (b) energy scores with Optiverse™ compounds, (c) contact scores using the Maybridge data base, and (d) energy scores using the Maybridge data base.

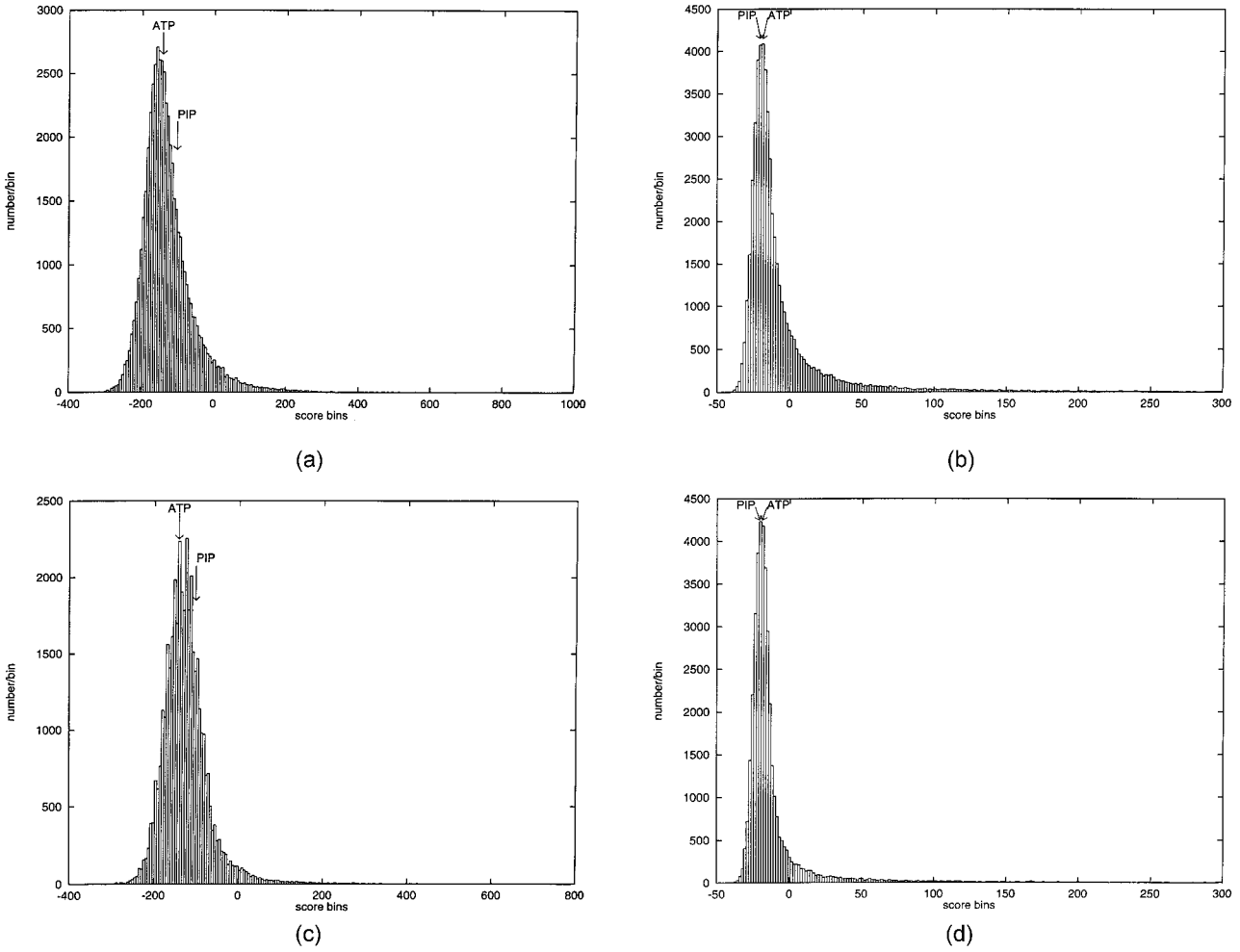

FIGURE 4. Histograms of PK docking scores. Arrows indicate the score of ligands in the X-ray conformation. (a) Contact scores with Optiverse^{TM} compounds, (b) energy scores with Optiverse^{TM} compounds, (c) contact scores using the Maybridge data base, and (d) energy scores using the Maybridge data base.

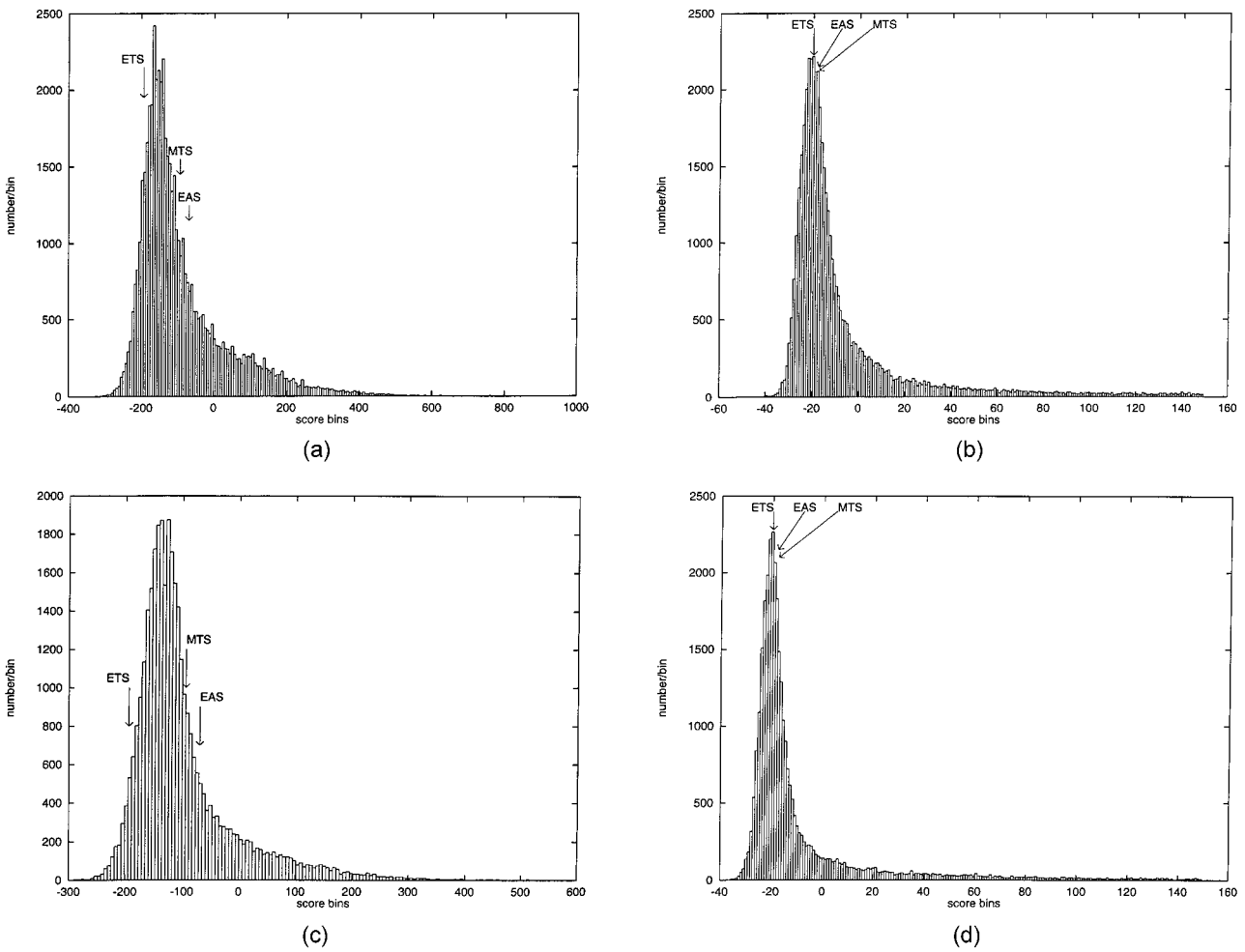

FIGURE 5. Histograms of CA docking scores. Arrows indicate the score of inhibitors in X-ray conformation. (a) Contact scores with Optiverse™ compounds, (b) energy scores with Optiverse™ compounds, (c) contact scores using the Maybridge data base, and (d) energy scores using the Maybridge data base.

TABLE II.
Statistical Analysis of Docking to LK, PK, and CA.

Docked Set	Mean	Mode	SD	Skew
LK Optiverse				
Contact	−102.49	−99.76	15.36	−0.79
Energy	−20.31	−20.36	2.63	0.04
LK Maybridge				
Contact	−93.20	−90.99	18.08	−0.45
Energy	−18.01	−18.79	3.19	0.86
LK* Maybridge				
Contact	−89.50	−96.98	50.02	2.79
Energy	−11.29	−12.98	18.88	8.92
PK Optiverse				
Contact	−122.17	−152.86	83.11	3.11
Energy	−4.81	−17.52	38.15	4.39
PK Maybridge				
Contact	−120.58	−119.78	57.75	2.37
Energy	−10.08	−18.65	31.27	5.55
PK* Optiverse				
Contact	−110.28	−141.27	89.39	3.21
Energy	−2.00	−18.24	43.67	4.05
CA Optiverse				
Contact	−83.78	−159.80	148.34	2.45
Energy	1.92	−18.94	55.19	3.67
CA Maybridge				
Contact	−95.05	−121.10	96.40	2.82
Energy	−2.44	−19.53	48.97	4.07

Docked set describes the combination of target and compound data bases, and energy and contact are the scoring functions. LK and PK denote the larger docking site while LK* and PK* indicate the smaller site consisting of a reduced number of spheres (see text). Mean is the sample average ($N - 1$); Mode is the most frequently observed value; SD is the standard deviation; and Skew is the skewness of the Gaussian distribution.

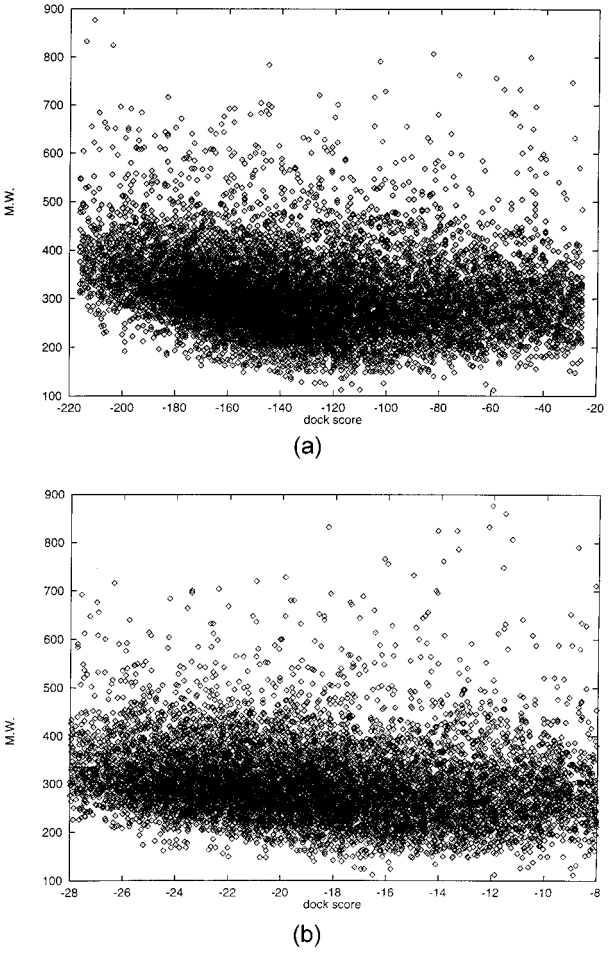

FIGURE 6. Molecular weight (MW) versus score plots for the peak region (within 0.5 standard deviation of the mode) of the Optiverse™ versus LK histograms. (a) MW versus contact score, (b) MW versus energy score.

sphere models differs (see above). The larger site in LK can accommodate more potential binding conformations than the smaller PK site. This is reflected by the broader peak of the LK distribution. The volume of the CA site is even smaller than in PK, increasing positive skewness of the distribution.

We further reduced the volume of the LK and PK sites by approximately 50% and redocked Maybridge and Optiverse™ compounds, respectively. The resulting distributions are more positively skewed than the corresponding distributions for the larger docking sites. Table III summarizes the volumes of all docking sites and the resulting skewness of the distributions. Docking into the smaller sites shifts the scores for many compounds above the peak level of the distributions.

In summary, our analysis shows that docking scores are not randomly distributed but are Gaussian-like. The skewness of the distributions depends on the volume of the docking sites. Given the imperfections of current scoring functions, a minimal docking site results in many compounds with scores significantly worse than the central mass.

Acknowledgment

We thank Nicola H. Chapman for helpful discussions.

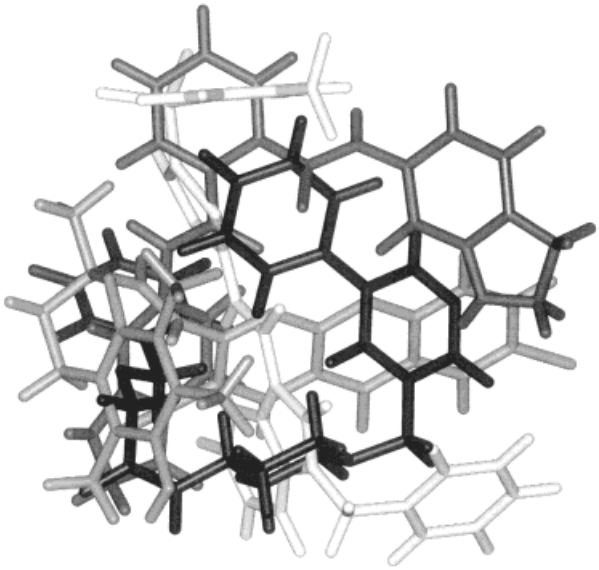

FIGURE 7. Binding mode of docked compounds from the peak region (within 0.5 standard deviation of the mode) of the Optiverse™ versus LK histogram. Different compounds are shown in gray scale. The LK docking site is omitted for clarity.

TABLE III.
Summary of Volumes of All Docking Sites.

	N	⟨Radius⟩	⟨Volume⟩	Total Volume	⟨Skew⟩
PK	62	2.04	40.60	664	2.74
PK*	24	2.06	42.36	456	3.21
LK	61	2.58	92.99	1375	−0.62
LK*	15	2.75	104.28	734	1.72
CA	32	2.00	43.11	500	2.63

N is the number of spheres in the DOCK model, ⟨Radius⟩ is the average radius of the spheres Å, ⟨Volume⟩ is the average volume of the spheres (Å^3), total volume is the approximate volume of the docking site Å^3, and ⟨Skew⟩ is the average contact score skewness.

References

1. Burt, S. K.; Hutchins, C. W.; Greer, J. Curr Opin Struct Biol 1991, 1, 213.
2. Cherfils, J.; Janin, J. Curr Opin Struct Biol 1993, 3, 265.
3. Blaney, J. M.; Dixon, J. S. Perspect Drug Discov Des 1993, 1, 301.
4. Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. Acc Chem Res 1994, 27, 117.
5. Lybrand, T. P. Curr Opin Struct Biol 1995, 5, 224.
6. Lengauer, T.; Rarey, M. Curr Opin Struct Biol 1996, 6, 402.
7. Kuntz, I. D. Science 1992, 257, 1078.
8. Greer, J.; Bush, B. L. Proc Natl Acad Sci USA 1978, 75, 303.
9. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Landgridge, R.; Ferrin, T. E. J Mol Biol 1982, 161, 269.
10. Leach, A. R.; Kuntz, I. D. J Comput Chem 1992, 13, 730.
11. Lorber, D. M.; Shoichet, B. K. Protein Sci 1998, 7, 938.
12. Leach, A. R. J Mol Biol 1994, 235, 345.
13. Ring, C. S.; Sun, E.; McKerrow, J. H.; Lee, G. K.; Rosenthal, P. J.; Kuntz, I. D.; Cohen, F. E. Proc Natl Acad Sci USA 1993, 90, 3583.
14. Shoichet, B. K.; Stroud, R. M.; Santi, D. V.; Kuntz, I. D.; Perry, K. M. Science 1993, 259, 1445.
15. Makino, S.; Kuntz, I. D. J Comput Chem 1997, 18, 1812.
16. Meng, E. C.; Gschwend, D. A.; Blaney, J. M.; Kuntz, I. D. Proteins Struct Funct Genet 1993, 17, 266.
17. Scolnick, L. R.; Clements, A. M.; Liao, J.; Crenshaw, L.; Hellberg, M.; May, J.; Dean, T. R.; Christianson, D. W. J Am Chem Soc 1997, 119, 850.
18. Engh, R. A.; Girod, A.; Kinzel, V.; Hubert, R.; Bossemeyer, D. J Biol Chem 1996, 271, 26157.
19. Yamaguchi, H.; Hendrickson, W. A. Nature 1996, 384, 484.
20. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J Mol Biol 1977, 112, 535.
21. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. J Comput Chem 1986, 7, 230.
22. Connolly, M. L. J Appl Crystallogr 1983, 16, 548.
23. Zheng, J.; Knighton, D. R.; Ten Eyck, L. F.; Karlsson, R.; Xuong, N.; Taylor, S. S.; Sowadski, J. M. Biochemistry 1993, 32, 2154.
24. Garr, C. D.; Peterson, J. R.; Schultz, L.; Oliver, A. R.; Underiner, T. L.; Cramer, R. D.; Ferguson, A. M.; Lawless, A. S.; Patterson, D. E. J. Biomol Screening 1996, 1, 179.
25. Gasteiger, J.; Marsili, M. Tetrahedron 1980, 36, 3210.
26. Böhm, H. J. J. Comput Aided Mol Des 1994, 8, 243.